2009 年度 博士論文要旨

Web 情報検索を支援する自然言語文脈処理に関する研究

関西学院大学大学院理工学研究科
情報専攻 北村研究室 高野 敦子

インターネットの普及に伴い，爆発的に増加している情報や知識を最大限に活用してユーザの知的な活動を支援する仕組みが求められている．それらの仕組みを実現するための技術として期待が寄せられている自然言語処理技術の中から，本論文では，インターネットとユーザとの間の対話処理及び Web 上で発信された評判情報の抽出を取り上げて論じる．

現在インターネットにおける高度な検索方法として期待されているのが，ユーザとコンピュータが自然言語を用いて対話を行いながら適切なサイトに誘導する対話型検索である．さらに，直接ユーザが必要とする情報をネット上から探し出して提示する研究も進められている．このような機能を実現するためには，ユーザとコンピュータが協調的にやり取りを行う過程で，最初は漠然としていたユーザの要求を明確化していく仕組みが必要となる．

一方，CGM（Consumer Generated Media）に代表されるように，個人がネット上に発信する意見や評判情報の重要性が広く認知されてきた．それらを有効利用するため，「口コミ情報」の中から評判情報を自動的に抽出・解析するための技術への関心が高まっている．

対話処理及び評判情報抽出はともに，汎用的な辞書情報以外に対象とする分野の領域知識を必要とする．最近の急速な知識の電子化と計算機の能力の向上により，領域知識の本格的な利用のための条件は整ってきた．しかし，分野や作業の種類に応じて効果的に領域知識を選定して導入する仕組みが実現されなければ，むしろ使用可能な知識量が増大するに従い不適正な意味処理を行う可能性が高まる．そのためには，領域知識を組み入れる前に，そのような知識の使用を前提としない核となる枠組みの構築とその限界の検証が必要であると考える．

自然言語処理は一般的に，形態素解析，構文解析，意味解析，文脈解析という順に処理が進み，次第に深い理解に基づく処理が行われる．そのうち，現在，形態素解析及び構文解析は汎用的な言語知識のみを用いても，ある程度高い精度の結果が得られる．しかし，その後続処理である意味解析を高い精度で行うためには領域知識が不可欠となる．それに対して筆者は，意味解析に続く文脈処理の一部は，精度の高い意味解析を必ずしも前提とせず，さらに対象分野に依存しない各言語活動固有の文脈構造を定式化することによって，領域知識を用いずに実現可能であると考える．本研究では，情報検索支援対話及び口コミ情報特有の構造を定式化することによって，領域知識を用いずに，不完全な意味解析あるいは構文解析に続く文脈処理を実現する枠組みを提案する．さらに，その枠組みを用いて具体的な文脈処理システムを構築することにより，その有効性と限界を検証す

る．

本来，対話が領域や作業の種類に独立に保持する特性として，一貫性と結束性という二つの性質が挙げられる．前者は対話内容の大域的な整合性や合目的性を意味し，後者は局所的なあるいは隣接する発話間の関連性を意味する．本研究では，局所的なやりとりにおける結束性認識を基礎とし，その上部構造として大域的に対話文脈を制御する一貫性を管理するという二重管理を導入する．これは，試行錯誤から生じる局所的なやりとりの多様性を結束性の認識によって吸収しようとする試みである．また本研究では，対話に普遍的な構造は表層構造にその解析のための手がかりが現れるとの考えから，表層表現から解析可能な対話の文脈構造を定式化し，限定された領域独立知識を用いた浅い文脈処理による対話管理手法を提案する．さらに，提案する手法の有効性を検証するために，局所的な発話間の結束性を質問とそれに対する応答に絞り，応答タイプを13種類に分類して形式化した．それに基づいて，応答タイプを絞り込む手法を提案する．

一方，掲示板などのロコミWebサイトにおける評判情報を自動で抽出する際には，「満足だ」，「悪い」といった評価を表す評価表現が重要な手がかりとなる．既に大規模な汎用的評価表現辞書は構築されているが，実際に評価を表すか否か，また好評か不評かは，評価する対象や観点に依存する．それに対して，本研究では，評価とそれに対する理由などの因果関係を持つ表現との間の文脈構造に着目した．「評価を表す表現は理由などの具体的な評価の内容を同時に示すことが多い」という仮説の元に，評価とそれに対して因果関係を持つ表現との間の表層構造を定式化した．それを用いて，汎用的な少数の評価表現を種として，ブートストラップ的に分野に特徴的な評価表現を自動で収集する仕組みを提案する．